PROCEEDINGS OF TDWG

Biodiversity Information Science and Standards 1: e20367
doi: 10.3897/tdwgproceedings.1.20367

OPEN ACCESS

Conference Abstract

A Comprehensive and Standards-Aware Common Data Model (CDM) for Taxonomic Research

Andreas Müller‡, Walter G. Berendsohn‡, Andreas Kohlbecker‡, Anton Güntsch‡, Patrick Plitzner‡, Katja Luther‡

‡ Botanic Garden and Botanical Museum, Freie Universität, Berlin, Germany

Corresponding author: Andreas Müller (a.mueller@bgbm.org), Walter G. Berendsohn (w.berendsohn@bgbm.org), Andreas Kohlbecker (a.kohlbecker@bgbm.org), Anton Güntsch (a.guentsch@bgbm.org)

Received: 16 Aug 2017 | Published: 16 Aug 2017

Citation: Müller A, Berendsohn W, Kohlbecker A, Güntsch A, Plitzner P, Luther K (2017) A Comprehensive and Standards-Aware Common Data Model (CDM) for Taxonomic Research. Proceedings of TDWG 1: e20367. https://doi.org/10.3897/tdwgproceedings.1.20367

Abstract

The EDIT Common Data Model (CDM) (FUB, BGBM 2008) is the centrepiece of the EDIT Platform for Cybertaxonomy (FUB, BGBM 2011, Ciardelli et al. 2009). Building on modelling efforts reaching back to the 1990ies, it aims to combine existing standards relevant to the taxonomic domain (but often designed for data exchange) with requirements of modern taxonomic tools. Modelled in the Unified Modelling Language (UML) (Booch et al. 2005), it offers an object oriented view on the information domain managed by expert taxonomists that is implemented independent of the used operating system and database management system (DBMS).

Being used in various national and international research projects with diverse foci over the past decade, the model evolved and became the common base of a variety of taxonomic projects, such as floras, faunas and checklists (see FUB, BGBM 2016 for a number of data portals created and made publicly available by different projects).

The CDM is strictly oriented towards the needs of the taxonomic experts community. Where requirements are complex it tries to reflect them reasonably rather than introducing ambiguity or reduced functionality via (over-)simplification. Where simplification is possible it tries to stay or become simple. Simplification on the model level is achieved by

© Müller A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

implementing business rules via constraints rather than via typification and subclassing. Simplification on the user interface level is achieved by numerous options for customisation.

Being used as a generic model for a variety of application types and use cases, it is adaptable and extendable by users and developers. It uses a combination of static and dynamic typification to allow both efficient handling of complex but well-defined data domains such as taxonomic classifications and nomenclature as well as less well-defined flexible domains like factual and descriptive data. Additionally it allows the creation of more than 30 types of user defined vocabularies such as those for taxonomic rank, nomenclatural status, name-to-name relationships, geographic area, presence status, etc.

A strong focus is set on good scientific praxis by making the source of almost all data citable in detail and offering data lineage to trace data back to its roots. It is also easy to reflect multiple opinions in parallel, e.g. differing taxonomic concepts (Berendsohn 1995, Berendsohn & al., this session) or several descriptive treatments obtained from different regional floras or faunas.

The CDM attempts to comprehensively cover the data used in the taxonomic domain - nomenclature, taxonomy (including concepts), taxon distribution data, descriptive data of all kinds, including morphological data referring to taxa and/or specimens, images and multimedia data of various kinds, and a complex system covering specimens and specimen derivatives down to DNA samples and sequences (Kilian et al. 2015, Stöver and Müller 2015) that mirrors the complexity of knowledge accumulation in the taxonomic research process.

In the context of the EDIT Platform, several applications have been developed based on the CDM and the library that provides the API and web Service interfaces based on the CDM (see Kohlbecker & al. and Güntsch & al., this session). In some areas the CDM is still evolving - although the basic structures are present, questions of application development feed back into modelling decisions. However, a "no-shortcuts" approach to modelling has variously delayed application development in the past, but it now pays off: the Platform can rapidly adapt to changing requirements from different projects and taxonomic specialists.

Keywords

EDIT Platform, Taxonomy, Modelling

Presenting author

Andreas Müller

Hosting institution

Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Germany

References

- Berendsohn W (1995) The Concept of "Potential Taxa" in Databases. Taxon 44 (2): 207-212. https://doi.org/10.2307/1222443
- Booch G, Rumbaugh J, Jacobson I (2005) Unified Modeling Language User Guide. 2nd Edition. Addison-Wesley, 496 pp. [In English]. [ISBN 0-321-26797-4]
- Ciardelli P, Kelbert P, Kohlbecker A, Hoffmann N, Güntsch A, Berendsohn WG (2009) The EDIT Platform for Cybertaxonomy and the Taxonomic Workflow: Selected Components. In: Fischer S, Maehle E, Reischuk R (Eds) INFORMATIK 2009, Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI). INFORMATIK 2009, Lübeck, 28.9. - 2.10.2009. Springer Lecture Notes in Informatics (LNI), 154, 625-638 pp. [In English].
- FUB, BGBM (2008) EDIT Common Data Model. http://cybertaxonomy.eu/cdm-uml. Accessed on: 2017-8-07.
- FUB, BGBM (2011) EDIT Platform for Cybertaxonomy. http://www.cybertaxonomy.org. Accessed on: 2017-8-07.
- FUB, BGBM (2016) EDIT Platform for Cybertaxonomy - Reference projects. https://cybertaxonomy.eu/?q=DataPortalReference. Accessed on: 2017-8-14.
- Kilian N, Henning T, Plitzner P, Müller A, Güntsch A, Stöver BC, Müller KF, Berendsohn WG, Borsch T (2015) Sample data processing in an additive and reproducible taxonomic workflow by using character data persistently linked to preserved individual specimens. *Database: The Journal of Biological Databases and Curation* 2015: 1-19. https://doi.org/10.1093/database/bav094
- Stöver BC, Müller KF (2015) LibrAlign - A powerful Java GUI library for MSA and attached raw and meta data. http://www2.ieb.uni-muenster.de/EvolBiodivPlants/en/Publications/ConferenceContribution?id=100872. Accessed on: 2017-8-15.